Toxicology and Applied Pharmacology 319 (2017) 39–46

Contents lists available at ScienceDirect

Toxicology and Applied Pharmacology

journal homepage: www.elsevier.com/locate/taap

High Impact Short Article

Improving the International Agency for Research on Cancer's consideration of mechanistic evidence

Julie Goodman *, Heather Lynch

Gradient, 20 University Road, Cambridge, MA 02138, United States

ARTICLE INFO

Article history:
Received 20 September 2016
Revised 10 January 2017
Accepted 27 January 2017
Available online 3 February 2017

Keywords:
IARC
Mode of Action
Hazard assessment
Carcinogen classification
Systematic review

ABSTRACT

Background: The International Agency for Research on Cancer (IARC) recently developed a framework for evaluating mechanistic evidence that includes a list of 10 key characteristics of carcinogens. This framework is useful for identifying and organizing large bodies of literature on carcinogenic mechanisms, but it lacks sufficient guidance for conducting evaluations that fully integrate mechanistic evidence into hazard assessments.
Objectives: We summarize the framework, and suggest approaches to strengthen the evaluation of mechanistic evidence using this framework.
Discussion: While the framework is useful for organizing mechanistic evidence, its lack of guidance for implementation limits its utility for understanding human carcinogenic potential. Specifically, it does not include explicit guidance for evaluating the biological significance of mechanistic endpoints, inter- and intra-individual variability, or study quality and relevance. It also does not explicitly address how mechanistic evidence should be integrated with other realms of evidence. Because mechanistic evidence is critical to understanding human cancer hazards, we recommend that IARC develop transparent and systematic guidelines for the use of this framework so that mechanistic evidence will be evaluated and integrated in a robust manner, and concurrently with other realms of evidence, to reach a final human cancer hazard conclusion.
Conclusions: IARC does not currently provide a standardized approach to evaluating mechanistic evidence. Incorporating the recommendations discussed here will make IARC analyses of mechanistic evidence more transparent, and lead to assessments of cancer hazards that reflect the weight of the scientific evidence and allow for scientifically defensible decision-making.

1. Introduction

The International Agency for Research on Cancer (IARC) published its first *Monographs on the Evaluation of Carcinogenic Risks to Humans* in 1971. Since that time, more than 900 agents have been evaluated for human carcinogenic potential (IARC, International Agency for Research on Cancer, 2015a). The monograph is intended to be a hazard evaluation, i.e., the first step in risk assessment. That is, IARC's goal is to identify whether a substance is associated with the development of cancer, regardless of the dose or exposure level at which an increased risk may occur. As a result, IARC states explicitly that it may identify an agent as a cancer hazard even when risks are very low at the exposure levels in the population of interest (IARC, International Agency for Research on Cancer, 2015b).

Each IARC monograph is written by a Working Group comprised of experts selected on the basis of knowledge and experience and the absence of "real or apparent conflicts of interest" (IARC, International Agency for Research on Cancer, 2015b). Invited experts with critical knowledge who have potential conflicts of interest may also be brought in to assist the Working Group and draft text on non-influential issues. The general roles of the Working Group members are outlined in the Preamble and Author Instructions. The Preamble summarizes scientific principles that govern the IARC *Monographs*; and the Author Instructions, which are intended to be used along with the Preamble, provide additional specifications to members of the Working Group writing the IARC monograph (IARC, International Agency for Research on Cancer, 2015b, 2016a). The instructions provide guidance on the literature search process, the organization of search results, the level of detail required for study summaries, the information to be provided in tables, and some brief considerations regarding animal and epidemiology study quality. Neither of these documents provide a step-by-step framework for reviewing studies, assessing quality, and integrating the evidence within or across each discipline.

While the Preamble and Author Instructions provide a general guide to the monograph evaluation process, the specific methodology varies by *Monograph*. In addition to the general Author Instructions, IARC provides monograph-specific instructions to the Working Group; these documents are not released publicly. Further, IARC explicitly states that, while the Preamble provides the overarching principles of the

* Corresponding author.
E-mail address: jgoodman@gradientcorp.com (J. Goodman).

http://dx.doi.org/10.1016/j.taap.2017.01.020

process, "The procedures through which a Working Group implements these principles are not specified in detail. They usually involve operations that have been established as being effective during previous monograph meetings but remain, predominantly, the prerogative of each individual Working Group" (IARC, International Agency for Research on Cancer, 2015b).

The main charge of the Working Group is to determine how an agent or group of agents should be classified within the IARC carcinogen classification framework (Table 1). IARC specifies that the categorization is a matter of scientific judgment that reflects the strength of evidence across the realms of evidence (IARC, International Agency for Research on Cancer, 2015b). Information on exposure levels in workers and the general population is summarized, but not generally factored into causal classifications, because IARC evaluates hazard and not risk. In some cases, dose-response data are summarized in an evaluation, though IARC provides no explicit guidance with regard to how these data should be interpreted in the context of causal conclusions (IARC, International Agency for Research on Cancer, 2015b). To arrive at a classification, Working Groups have historically focused their reviews on epidemiology and animal bioassays deemed relevant and appropriate. Although they have also considered available mechanistic evidence, this was generally considered secondary to other realms of evidence.

There has been a recent shift in focus at IARC, whereby mechanistic evidence is given more weight in cancer hazard evaluations. As discussed below, the IARC framework is useful for organizing mechanistic evidence, but its lack of guidance for implementation limits its utility for understanding human carcinogenic potential. Specifically, it does not include explicit guidance for evaluating the biological significance of mechanistic endpoints, inter- and intra-individual variability, or study quality and relevance. It also does not explicitly address how mechanistic evidence should be integrated with other realms of evidence. Because mechanistic evidence is critical to understanding human cancer hazards, we recommend that IARC develop transparent and systematic guidelines for the use of this framework, so that mechanistic evidence will be evaluated and integrated in a robust manner, and concurrently with other realms of evidence. This will result in assessments that are based on the best available science and, thus, allow for more scientifically defensible decision-making.

Table 1
IARC carcinogenicity classification system[a].

Classification	Requirements
Group 1 Carcinogenic to humans	▪ Sufficient evidence in humans OR ▪ Exceptionally, sufficient evidence in animals AND **strong evidence in exposed humans that the agent acts through a relevant mechanism** OR ▪ Clearly belongs, **based on mechanistic considerations**, to a class of agents for which one or more members have been classified in Group 1[b]
Group 2A Probably carcinogenic to humans	▪ Limited in humans AND sufficient in animals ▪ Inadequate in humans AND sufficient in animals AND **strong evidence that carcinogenesis is mediated by a mechanism that also operates in humans** ▪ Exceptionally, an agent may be classified in this category solely on the basis of limited evidence in humans[c] ▪ Clearly belongs, **based on mechanistic considerations**, to a class of agents for which one or more members have been classified in Group 2A
Group 2B Possibly carcinogenic to humans	▪ Limited in humans AND less than sufficient in animals ▪ Inadequate in humans BUT sufficient in animals ▪ Inadequate in humans AND less than sufficient in animals AND **supporting evidence from mechanistic and other relevant data** ▪ An agent may be classified in this category solely on the basis of strong evidence from mechanistic and other relevant data.[c]
Group 3 Not classifiable as to its carcinogenicity in humans	▪ Inadequate in humans AND inadequate/-limited in animals. ▪ Inadequate in humans AND sufficient in animals AND **strong evidence that the mechanism of carcinogenicity in animals does not operate in humans.**
Group 4 Probably not carcinogenic to humans[d]	▪ Sufficient evidence suggesting lack of carcinogenicity in humans and animals ▪ In some instances, inadequate evidence of carcinogenicity BUT evidence suggesting lack of carcinogenicity in experimental animals, **supporting evidence from mechanistic and other relevant data**

[a] As presented in the Preamble (IARC 2015b) and Author Instructions (IARC 2016a).
[b] Does not appear in the Preamble.
[c] This is only noted in the Preamble; it does not appear in the Author Instructions.
[d] The requirements for this category are not discussed in the Author Instructions.

2. IARC's use of mechanistic data

2.1. Overall approach according to the Preamble

The goal of the monographs has historically been to determine cancer hazard regardless of underlying mechanism; however, the current Preamble (IARC, International Agency for Research on Cancer, 2015b) and associated guidance materials have shifted focus to include information on mechanisms in the overall evaluation of an agent. The Preamble specifies that the Working Group is charged with identifying possible mechanisms whereby an agent of interest may increase the risk of cancer, and when available, summarize a representative "selection of key mechanistic data"; the Preamble explicitly states that a monograph need not cite all mechanistic literature for the agent, but does not give direction on how to identify key studies (IARC, International Agency for Research on Cancer, 2015b).

In the Preamble, mechanisms are grouped into physiological changes (e.g., escape from apoptosis and/or senescence), functional changes at the cellular level (e.g., changes in gene expression), and changes at the molecular level (e.g., DNA adducts and DNA strand breaks). Mechanistic data are discussed in their own section of the monograph, and then considered within the overall Evaluation and Rationale as they relate to plausibility of effects observed in animals. The strength of evidence that any observed carcinogenic effect in animals is due to a specific mechanism is rated as "weak," "moderate," or "strong." Evidence that a mechanism operates in animals is strengthened if results are consistent in different species; data are coherent; and studies show that when the relevant mechanism is suppressed, tumor development is also suppressed (IARC, International Agency for Research on Cancer, 2015b). Specific guidelines for ranking the strength of mechanistic evidence are not detailed, however, and there is no discussion of what actually constitutes "weak," "moderate," or "strong" mechanistic evidence.

After reaching conclusions on the strength of mechanistic evidence, IARC determines whether a particular mechanism is likely to operate in humans; most often, this conclusion is made if there are measured data in humans or biological specimens from humans.

2.2. 10 key mechanism-of-action characteristics

When IARC reviewed Group 1 carcinogens in *Monograph* 100 in early 2011, it noted that many were classified before mechanistic data were available, and that these data had become available in the prior two decades. IARC also found that the agents it had listed as human carcinogens shared a number of common characteristics (Smith et al., 2016). In 2012, IARC organized two workshops to discuss the mechanisms by which Group 1 carcinogens cause cancer; the participants

concluded that these agents typically exhibit one or more of the 10 "key" characteristics (See Table 2). These characteristics were informed by Hanahan and Weinberg's "hallmarks of cancer," which are characteristics common to cancer cells (e.g., the ability for cancer cells to avoid cell senescence) (Hanahan and Weinberg, 2011). Many of IARC's 10 characteristics are those whereby chemicals can directly initiate carcinogenic processes in cells, or enable an environment favorable to tumor formation (e.g., disrupting cellular pathways to facilitate aberrant replication by cancer cells).

After the 10 characteristics were identified, IARC participants developed a framework for identifying and organizing mechanistic data around these characteristics, and published this framework in *Environmental Health Perspectives* (Smith et al., 2016). Discussion of the 10 key characteristics only appears in the 2016 Author Instructions and in Smith et al. (2016); there is no explicit discussion of this framework in the current version of the Preamble. The characteristics first appeared in monographs released in 2015.

The key characteristics framework may be a useful way to initially categorize mechanistic evidence, but it fails to describe how the quality, external validity, and relevance of the evidence should be considered, or how positive and negative findings should be integrated to make conclusions on the likelihood that a substance has any given mechanism, or causes cancer through that mechanism. Further, IARC fails to fully consider that some of the 10 key characteristics are also shared by non-carcinogenic agents in its framework (although this is acknowledged by Smith et al., 2016). This sets a problematic precedent for agents possessing a specific characteristic, or falling into a class of similar substances. This is because, no matter how strong the epidemiology and toxicity evidence suggesting otherwise, there is a possibility that the agent will be assumed to have some carcinogenic hazard based on mechanistic evidence alone. This will certainly not always be the case, but there is nothing in the framework to prevent it from occurring.

Table 2
IARC's 10 key characteristics of carcinogens[a,b].

Characteristic	Examples of relevant evidence
1) Is electrophilic or can be metabolically activated	Parent compound or metabolite with an electrophilic structure (e.g., epoxide, quinone, etc.), formation of DNA and protein adducts
2) Is genotoxic	DNA damage (DNA strand breaks, DNA-protein cross-links, unscheduled DNA synthesis), intercalation, gene mutations, cytogenetic changes (e.g., chromosome aberrations, micronuclei)
3) Alters DNA repair or causes genomic instability	Alterations of DNA replication or repair (e.g., topoisomerase II, base-excision or double-strand break repair)
4) Induces epigenetic alterations	DNA methylation, histone modification, microRNA expression
5) Induces oxidative stress	Oxygen radicals, oxidative stress, oxidative damage to macromolecules (e.g., DNA, lipids).
6) Induces chronic inflammation	Elevated white blood cells, myeloperoxidase activity, altered cytokine and/or chemokine production.
7) Is immunosuppressive	Decreased immunosurveillance, immune system dysfunction.
8) Modulates receptor--mediated effects	Receptor in/activation (e.g., ER, PPAR, AhR) or modulation of exogenous ligands (including hormones).
9) Causes immortalization	Inhibition of senescence, cell transformation.
10) Alters cell proliferation, cell death, or nutrient supply	Increased proliferation, decreased apoptosis, changes in growth factors, energetics and signaling pathways related to cellular replication or cell cycle control, angiogenesis.

Notes: AhR = Aryl Hydrocarbon Receptor; DNA = Deoxyribonucleic Acid; ER = Estrogen Receptor; PPAR = Peroxisome Proliferator-activated Receptor; RNA = Ribonucleic Acid.

[a] Any of the 10 characteristics in this table could interact with any other (e.g., oxidative stress, DNA damage, and chronic inflammation combined provide stronger evidence for a cancer mechanism than would oxidative stress alone).

[b] Adapted from Smith et al. (2016).

As noted by Smith et al. (2016), information organized by the 10 characteristics can be used to develop a Mode of Action (MOA) or adverse outcome pathway (AOP), but there is a substantial difference between listing potential mechanistic endpoints involved in the well-established carcinogenicity of benzene (discussed on p 718 of Smith et al., 2016) and the integration of this information with human and animal evidence to support a cancer classification, as presented in IARC volumes 112 and 113.

Overall, IARC's current "10 key characteristics" framework for evaluating mechanistic evidence is incomplete and fails to fully and appropriately integrate mechanistic evidence into carcinogen hazard classifications. Like epidemiology and toxicity evidence, mechanistic evidence should be assessed for quality and relevance to determine whether it supports a biologically plausible path to cancer, and not be given more weight than other types of evidence if it is not of similar quality and relevance.

3. Concerns with application of the key characteristics

3.1. Key characteristics do not always indicate a cancer hazard

IARC identified its 10 key characteristics based on the processes exhibited by cancerous cells. Because some of these characteristics are common to agents that do not cause cancer, the approach is akin to case reports in epidemiology. In case reports (or case series), one looks for similarities among cases, but there are no healthy control populations with which to differentiate the cases from non-affected people. While case reports may be useful for identifying areas for future research, or for generating hypotheses, they cannot be used alone to assess causation. Similarly, while the 10 key characteristics may be used to organize and evaluate possible mechanisms of action for carcinogenic agents, the existing IARC approach should not be used in isolation to support or refute the biological plausibility of any postulated carcinogenic mechanism.

In some cases, a substance may possess one or more key characteristics, but not be a carcinogen (Goodson et al., 2015). For example, the pesticide paraquat is known to be highly toxic via oxidative stress mechanisms (specifically, lipid peroxidation), but it does not cause cancer in animals or humans. While some in vitro micronucleus assays are positive, in its evaluation of paraquat, the Food and Agriculture Organization of the United Nations (FAO)/World Health Organization (WHO) Codex Committee on Pesticide Residues noted that the positive results were likely due solely to the high doses used in the assays. FAO/WHO specifically noted that the "well established" oxidative stress mechanism only manifests at high doses, owing to saturation of cellular defensive mechanisms, but that "it is likely there would be a threshold as, except at high doses, reactive oxygen species are rapidly detoxified" (FAO/WHO, 2004). In humans, the doses necessary to cause sustained oxidative damage would likely be of the magnitude associated with death from poisoning. Using IARC's framework, however, paraquat would be considered to have some degree of carcinogenic hazard.

Paraquat illustrates an issue with the current IARC approach for assessing mechanistic data; i.e., the Working Group staff may simply identify that an agent possesses a characteristic (or several) and conclude based on the presence of this mechanism that there is moderate or strong evidence of a carcinogenic mechanism.

3.2. Biological relevance

The 10 key characteristics framework does not discuss how biologically relevant specific endpoints or assays are to humans. As highlighted by Miller et al. (2016), it can be difficult to determine whether biological perturbations represent homeostatic changes or are molecular initiating events that may lead to cancer. This is most notable for genotoxicity, where different assays differ in their predictive ability with regard to cancer. IARC's genotoxicity category encompasses any characteristic

associated with DNA damage, mutation, or both, via direct DNA damage (e.g., DNA double strand breaks and 8-oxo-2′-deoxyguanosine [8-OHdG] lesions), mutagenicity (changes in the sequence of standard base pairs of DNA) and/or clastogenicity (i.e., damage that results in the addition, deletion, or rearrangement of sections of a chromosome). These characteristics differ in that some, such as DNA damage and 8-OHdG lesions, are repairable effects that may or may not be associated with downstream carcinogenic processes. In contrast, clastogenicity (e.g., chromosomal aberrations) are not repairable (although not all mutagenic or clastogenic agents are always carcinogenic). Also, there are numerous assays for each of these endpoints, and their precision and applicability to humans varies considerably (see Preston and Hoffman, 2013).

The 10 key characteristics framework also makes no mention of the predictive ability of in vitro assays relative to in vivo studies (Butterworth, 2006). There are numerous assays for each of these endpoints, and their precision and applicability and relevance to humans varies considerably (see Preston and Hoffman, 2013). Studies have reported a high false positive rate for in vitro genotoxicity and mutagenicity tests relative to rodent studies; therefore, in vitro assays should never be used alone for assessing cancer risk (Butterworth, 2006; Morita et al., 2009). Despite this, it does not appear that IARC gives less weight to in vitro assays.

Similarly, many studies targeting characteristics such as inflammatory potential are not designed to detect sustained changes (e.g., controlled human exposure studies of inflammatory markers in the lung or blood), nor determine the magnitude of change at the molecular or tissue level required to initiate a carcinogenic process (e.g., from change in biomarkers through cancer development). This makes it difficult to determine, using the available assays alone, whether observed effects persist over time and are sufficient to induce later carcinogenic processes. This is particularly problematic when assessing relevance to human exposure patterns, which typically vary over time (as is the case with many environmental exposures). For example, it is unclear what the amount of 8-OHdG, a common marker of DNA damage in animals and humans, is needed to overwhelm homeostatic repair mechanisms and facilitate the phenotype of cancer (Pilger and Rudiger, 2006).

Overall, the IARC 10 key characteristics framework does not distinguish between early and later endpoints associated with cancer development, and gives no guidance regarding which assays or endpoints associated with a given characteristic are more reliable and more likely associated with cancer risk. As a result, all characteristics and endpoints appear to be given equal weight in IARC evaluations, regardless of whether they are early or late on the pathway to cancer, and without regard for their relevance to human cancer.

3.3. Intra- and inter-individual variability

Another potential issue associated with many of the 10 key characteristics is that there is some difficulty in accurately measuring biological changes induced by an agent versus those caused by population differences, natural fluctuations, and other non-cancer infections and diseases.

It is well known that there is substantial inter-individual variability in "baseline" levels of biomarkers of effect. Factors that affect these measures include genetics, lifestyle choices (e.g., smoking, diet), age, and health history (e.g., chronic infection) (Au, 2007). For example, population-specific genetic variation can affect baseline levels of microRNA (miRNA), proteins involved in regulation of gene expression, commonly used in cancer biomarker studies. Rawlings-Goss et al. (2014) found that the baseline levels (i.e., measurements taken in the absence of apparent disease) of several circulating miRNAs associated with susceptibility to certain cancers were differently expressed in different global populations (European, Asian, and African).

In addition to differences within and across populations, fluctuations in baseline levels of biomarkers within an individual (intra-individual variability) may also affect interpretation of changes in various biomarkers. Many biomarkers may fluctuate on a daily basis (e.g., 8-OHdG); thus, short-term fluctuations in biomarkers over the course of a study may be caused by these fluctuations, as opposed to the agent under study.

In addition, several of these key characteristics may be associated with multiple disease states, both acute and chronic, including but not limited to cancer in an individual. In other words, observed changes may be the result of another disease state, and not indicative that cancer is likely to occur (Pilger et al., 2001). This means that studies measuring biomarkers could potentially attribute changes in biomarkers in an individual or group to a cancer-causing agent when, in fact, those changes may have been caused by another underlying condition or disease state. For example, many disease states are associated with chronic inflammation and DNA damage, such as diabetes, heart disease, and Alzheimer's disease. Although some study designs control for some sources of observed changes outside of the agent of interest, including intra- and inter-person variability (e.g., cross-over human studies in which the biomarker is measured before and after exposure in the same person), unmeasured/unknown disease states could be altering levels of oxidative damage markers over the course of a study, potentially with unstable (changing) activity (Pilger et al., 2001).

As another example, the influence of infectious disease may not be fully accounted for in mechanistic studies of inflammation in humans (and resulting downstream effects, such as DNA methylation); biomarker studies in humans typically do not collect information on acute infections in participants during the study period (Lynch et al., 2017). Most viral or bacterial infections are capable of inducing chronic inflammation, and in the case of acute infection, sometimes persisting well after the infection has cleared (Hattori and Ushijima, 2016; Li et al., 2015). As such, the ability to detect inflammation in mechanistic studies cannot be used as evidence, in isolation, to determine the likelihood of a carcinogenic mechanism from exposure to the agent of interest.

The IARC framework offers no suggestions for how Working Groups should address issues with respect to intra- and inter-individual variability in mechanistic data.

3.4. Study quality

Another important gap in the IARC evaluation process is the absence of explicit guidance for evaluating study quality. Although there are some aspects of study quality alluded to in the Preamble and Instructions to Authors, there is no explicit guidance on what methodological strengths and limitations to consider, and how these factors impact the interpretation of results in individual studies and when considering the body of literature as a whole. Numerous agencies, including the National Research Council, have demonstrated how quality assessment is a critical step in the systematic review process, and provide suggestions for how it can be done (Lynch et al., 2016; Stephens et al., 2016).

There are few study quality assessment frameworks that are tailored specifically for mechanistic studies (Samuel et al., 2016), but there are several that can be used with little adaptation to assess mechanistic data for both methodological and reporting quality, including the Klimisch System and the related ToxRTool. The Klimisch system (Klimisch et al., 1997) is based largely on the principals of Good Laboratory Practice (GLP) and OECD (Organization for Economic Development) test guidelines, and the ToxRTool (European Commission, 2017) is a newer, computer-based system that refines the general Klimisch principles (Schneider et al., 2009). Both systems provide a framework to evaluate studies based on quality considerations, including study size (or for in vitro assays, replicates), blinding, experimental procedure, outcome assessment methods, and an author's consideration of factors that may influence the interpretation of in vitro or other experimental assays (e.g., pH shift, impurities, test substance solubility). A Klimisch score of 1 ("reliable without restrictions") is

reserved for guideline studies (e.g., OECD) and those otherwise comparable to guideline studies. A score of 2 (reliable with restrictions) is for a non-guideline study that is of a similar quality to a guideline study, based on the categories mentioned above. A score of 3 ("not reliable") is given to non-guideline studies with methods that have not been validated, and studies that do not provide adequate documentation to assess validity. The criteria then inform an overall reliability classification ("reliable without restriction," "reliable with restrictions," "not reliable," or "not assignable") (Lynch et al., 2016). Recent weight-of-evidence reviews also provide examples of how existing frameworks can be tailored to fit mechanistic data and accommodate chemical-specific issues (e.g., Goodman et al., 2015; Lynch et al., 2017).

In short, there are many resources IARC could apply to ensure Working Groups are adequately considering study quality. Without specific guidance, each Working Group may interpret the evidence differently with regard to strengths and limitations. As a consequence, in several instances, evidence for a key characteristic could be mischaracterized as stronger or weaker than it actually is.

3.5. Integrating mechanistic evidence

The final major limitation of the current IARC framework is that it does not specify how results within and across each characteristic category should be integrated to determine the weight of evidence for any carcinogenic mechanism(s). In the footnote in Table 1 of the Author Instructions (IARC, International Agency for Research on Cancer, 2016a), IARC states, "Any of the 10 characteristics in this table could interact with any other (e.g., oxidative stress, DNA damage and chronic inflammation), which when combined provides stronger evidence for a cancer mechanism than would oxidative stress alone." There is no discussion, however, with regard to what exactly this means – i.e., what the presence of only a single characteristic means, and how specifically multiple characteristics may interact to strengthen the evidence of a potential carcinogenic mechanism.

Further, there is no discussion of how the presence of one and absence of another related characteristic may cast doubt on an agent's carcinogenic potential. For example, if an agent is able to cause DNA damage, but does not induce chromosomal aberrations or micronuclei, this indicates that there is insufficient evidence to support a mechanism of carcinogenicity. In contrast, if an agent causes chronic inflammation and exhibits strong evidence of other characteristics (e.g., it induces chromosomal aberrations), and inflammation occurs in studies that also report downstream effects (e.g., neoplastic lesions in animals), then chronic inflammation may be considered a key step in the MOA of that agent. Because the IARC framework does not require consideration of the characteristics in the context of one another, however, some IARC assessments may be lacking a complete assessment of the coherence of mechanistic evidence. IARC should follow an approach that assesses a complete mechanistic pathway, rather than a "single endpoint" or groups of potentially unrelated endpoints, as recommended by others evaluating cancer biomarkers for environmental chemicals (e.g., Goodson et al., 2015).

Whether an agent has any of the 10 key characteristics may be important when considered in the greater context of the weight of mechanistic evidence; however, the presence of any one characteristic in isolation cannot be used as support for the classification of an agent as a human carcinogen. The key characteristics framework would be more effective if Working Group members evaluated the weight of evidence within each characteristic, considering the factors outlined above, and then assessed how the evidence for each of the characteristics should be interpreted in the context of one another and epidemiology and toxicity evidence. In other words, IARC should provide explicit guidance on interpreting the mechanistic evidence with regard to biological relevance and assay reliability, intra- and inter-individual variability, and study quality both within and across the 10 key characteristics, and considering other realms of evidence. This approach will encourage IARC Working Group members to more fully evaluate the strengths and limitations of the available literature to arrive at an overall confidence rating for the body of mechanistic literature as a whole and in the context of other realms of evidence.

Finally, the evaluation of mechanistic evidence should be incorporated into a larger weight-of-evidence assessment that considers the mechanistic evidence equally and in tandem with the other realms of evidence for the purposes of final carcinogenic hazard classification. At present, the IARC framework is a strength-of-evidence approach; i.e., positive, statistically significant evidence from a single study may be considered sufficient for the basis of a causal classification (Willhite, 2001). The weight of the body of the mechanistic evidence should be considered, then integrated with the other realms of evidence (human and animal). If this is not done, the carcinogenic classification may be untenable.

4. Weight of evidence and human relevance

Many of the potential issues with interpretation of IARC's 10 key characteristics, and mechanistic evidence in general, can be addressed using a structured systematic framework across all IARC reviews. Smith et al. (2016) stated that at the time of the 2012 IARC review, there was no readily available "broadly accepted" systematic method for identifying and organizing mechanistic evidence. In fact, there were several frameworks for incorporating mechanistic evidence in hazard and risk assessment available prior to 2012, most notably, the International Program on Chemical Safety (IPCS) Mode of Action Human Relevance (MOA/HR) framework (Table 3). The IPCS framework was developed by a large, multi-disciplinary group of scientists and was derived from earlier work by US EPA (Boobis et al., 2008; Meek et al., 2014a). It has been applied successfully to more than 30 case studies and adopted by national and international agencies to assist in transparency and consistency in MOA assessments (Meek et al., 2014a). We note that other options have been published after 2012, including the AOP framework (Villeneuve et al., 2014).

Details of the MOA/HR framework and subsequent updates are available (see, for example, Boobis et al., 2008; Meek et al., 2003, 2014a, 2014b). Briefly, this framework provides guidance for evaluating the WoE for key events of an MOA using modified Bradford Hill considerations (Hill, 1965) and analyzing concordance of the key events within and among species (Meek et al., 2014a, 2014b). First, the hazard assessor is charged with establishing whether a plausible MOA exists in animals. After compiling studies and determining the postulated key events in the MOA, the assessor then considers the presence of dose-response relationships; temporal associations; strength,

Table 3
IPCS MOA/HR framework.
Adapted from Meek et al. (2003).

1) Is the weight of evidence sufficient to establish the MOA in animals?
   a) Postulated MOA
   b) Identification of key events
   c) Animal evidence
   d) Application of US EPA/IPCS animal MOA guidance
2) Are key events in the animal MOA plausible in humans?
   a) Concordance analysis of human and animal responses (qualitative and quantitative)
   b) Statement of confidence
3) Taking into account kinetic and dynamic factors, is the MOA plausible in humans?
   a) Concordance analysis of human and animal responses
   b) Statement of confidence
4) Statement of confidence; analysis; implications for risk assessment

Notes: HR = Human Relevance; IPCS = International Programme on Chemical Safety; MOA = Mode of Action; US EPA = United States Environmental Protection Agency.

consistency, and specificity; biological plausibility; alternative MOAs that may be applicable; and any uncertainties, inconsistencies, and data gaps in the mechanistic literature. If the weight of evidence suggests a MOA operates in animals, the assessor must then consider whether the MOA evidence is in concordance with the larger body of evidence. For example, the reviewer is asked to consider questions such as, "Are the key events observed at doses below or similar to those associated with the end (adverse) effect?," and "Is the pattern of observations across species/strains/organs/test systems what would be expected based on the hypothesized MOA?" If the WOE for the hypothesized MOA is sufficient and relevant to humans, implications for dose-response in humans are then considered with regard to pharmacokinetic and pharmacodynamic data.

The MOA/HR framework facilitates a thorough analysis of mechanistic evidence within a larger weight-of-evidence assessment to determine whether any observed MOA(s) plausibly operate in humans at relevant exposure levels. Updates to the MOA/HR framework have focused on the importance of an iterative approach and the application of the framework to substances with less measured data available (Meek et al., 2014a); as such, the framework continues to evolve to accommodate new developments in the field. The IARC 10 key characteristics framework is not as systematic, clear, and thorough as the MOA/HR framework. While individual Working Group staff may follow their own process that mirrors the MOA/HR framework, to ensure a consistent, unbiased review, the IARC Preamble and Author Instructions should maintain that all Working Group staff across all monographs follow the same, systematic process for evaluating mechanistic evidence. IARC could easily adapt the MOA/HR framework or something similar to aid in a unified approach across Working Groups.

5. Case study: diazinon

In 2016, IARC released a monograph that included an assessment of diazinon, an organophosphate insecticide used to control insects, nematodes, and mites (IARC, International Agency for Research on Cancer, 2016b). Diazinon was allowed for residential use until 2004, but current production and use is limited, even in agriculture. In its evaluation, the Working Group determined that while the animal and human data were limited, the strong evidence that diazinon can operate through two key characteristics of carcinogens (genotoxicity and oxidative stress) provided support for the classification that diazinon is *probably carcinogenic to humans* (Group 2A) (IARC, International Agency for Research on Cancer, 2015c, 2016b).

Diazinon was evaluated in the first monograph to organize mechanistic evidence by the 10 key characteristics. The Working Group concluded that the overall body of evidence suggested that diazinon may be genotoxic because it consistently caused DNA damage in animals in vivo and human cells in vitro (e.g., DNA strand breaks, increased sister chromatid exchange), and micronuclei formation, but the results of clastogenicity assays were mixed (specifically, the chromosomal aberration tests). The Working Group concluded that other characteristics have been less well studied and in some cases, have shown mixed results. Notably, human cell in vitro assays show virtually no effects on cell viability or proliferation; in some experiments, diazinon caused a *decrease* in cell growth at increasing doses. Similarly, findings from in vivo and in vitro animal assays of proliferation are limited and mixed, although they include some evidence of increased proliferation in different cell lines (e.g., rat pituitary tumor cells, neuronal and mixed cortical cells).

The diazinon monograph also discusses an analysis in which Working Group members mapped endpoints in the ToxCast/Tox21 database to the key characteristics, then determined if diazinon was "active" or "inactive" for each of the selected assay endpoints using ToxPi software. This software assigned the potential for diazinon to be associated with each key characteristic relative to 178 other chemicals that have been previously evaluated by the IARC monographs and screened by ToxCast. IARC reported that there were several "active" endpoints, including for characteristic 1 (is electrophilic): CYP inhibition, aromatase inhibition; characteristic 2 (is genotoxic): two endpoints of TP53 activity; and characteristic 8 (modulates receptor-mediated effects): 16 endpoints, including estrogen receptor alpha and beta, AhR; and three cytotoxicity endpoints. Diazinon was considered inactive for chronic inflammation and epigenetic modifications and had "negligible activity" for oxidative stress. ToxPi scores for diazinon were generally well below (generally 2–22 times lower) those of the highest-ranked chemicals in the database for each of the characteristic (and ranked toward the low end of the graphs with regard to activity for each endpoint). One must also consider the uncertainties associated with these high-throughput tests; recent analyses of ToxCast suggest that the program's results may not align with in vivo evidence of carcinogenic activity. For example, Cox et al. (2016) found that ToxCast was unable to accurately predict carcinogenic hazard in animal bioassays for most of the 292 non-genotoxic chemicals the authors assessed.

The diazinon monograph gives no indication of whether it captured the entire body of relevant mechanistic literature (including all aspects of relevant studies), or whether there was any consideration of study quality. The results of assays of genetic and "related effects" are summarized in the monograph's Tables 4.1 through 4.5, including a column with "comments"; however, the Working Group did not indicate which studies followed recognized guidelines (i.e., GLP, OECD), or provide any information about assay methods. It provides notations in the tables in cases where studies with positive results were of "limited quality"; however, nowhere in the monograph is there a discussion of how the Working Group determined overall quality, nor how quality may have factored into the interpretation of the mechanistic evidence. Further, there is no indication that the Working Group applied Bradford Hill considerations (or something similar) when assessing the body of mechanistic evidence, e.g., an assessment of dose-response relationships; temporal associations; strength, consistency, and specificity; biological plausibility; alternative mechanisms that may be applicable; and any uncertainties, inconsistencies, and data gaps. The monograph also did not discuss how null findings in certain mechanistic assays impacted the interpretation of positive findings in other assays. There was also no discussion on how mechanistic evidence was integrated with the other available evidence, i.e., how the (albeit limited) human and toxicity evidence impacted the interpretation of mechanistic evidence and vice versa. Regarding the ToxCast/Tox21 mapping, the monograph offers no explanation of how to interpret the output. It appears that the Working Group is insinuating by being "active" for a particular characteristic, or one of a class of chemicals with similar characteristics, this constitutes evidence for a plausible mechanism, and that mechanistic evidence can be considered strong or moderate based on a single "active" endpoint, and then used to upgrade a causal classification. As discussed in detail throughout this paper, merely falling into a class does not constitute evidence for causation; a much deeper analysis is needed.

It appears the Working Group's conclusion that diazinon is "probably carcinogenic to humans" is based heavily on the mechanistic evidence. While the database of available diazinon literature related to the 10 key characteristics is large, the Working Group did not fully consider factors that affect the interpretation of mechanistic endpoints, including biological relevance, variability, study quality, and human relevance. Applying the MOA/HR framework (described above) or something similar would have allowed for a more comprehensive evaluation of the mechanistic evidence to determine whether it is sufficient to support a specific hazard classification without sufficient epidemiology and animal evidence.

6. Conclusion

IARC's list of 10 key characteristics may help one identify and organize mechanistic evidence for evaluation. However, these characteristics are currently being used like a checklist, i.e., IARC guidance

suggests that simply having one or more characteristic is considered evidence of a carcinogenic mechanism. The framework does not provide a standardized approach for all Working Groups to determine how reliable and/or predictive specific endpoints are of cancer risk. In addition, there is no discussion of how to deal with inter- and intra-individual variability in biomarkers or how this may affect interpretation of the study results. Moreover, study quality and human relevance are not given sufficient consideration, and there is no indication of how the mechanistic data should be integrated in the assessment. The current IARC Preamble also suggests that mechanistic evidence should be used to up- or down-grade a casual classification based on human and animal evidence; as a result, it appears to be given an inappropriate weight relative to human and animal evidence. As illustrated with the case study of diazinon, the presence of one or more characteristics may result in a classification (e.g., probably carcinogenic) even in the absence of strong animal and/or human evidence.

Rather than focus on whether agents possess characteristics that are not necessarily specific to carcinogens, or fall into a certain class of chemicals, IARC should provide clear, explicit guidance for how to consider the totality of the mechanistic evidence, including study strengths and limitations, and how they impact the interpretation of results. This can be done by adapting available frameworks that address the issues of study quality and human relevance. Adopting a systematic approach for evaluating and integrating mechanistic evidence with the other realms of evidence will allow for hazard classifications that are scientifically defensible and appropriate for regulatory decision-making.

Transparency document

The Transparency document associated with this article can be found, in online version.

Conflict of interest statement

The authors are employed by Gradient, a private environmental consulting company. The authors have sole responsibility for the writing and contents of this paper. This paper represents the professional opinions of the authors and not necessarily those of the American Chemistry Council (ACC), which provided funding for this paper. The authors declare that they have no other actual or potential competing financial interests.

Competing financial interests declaration

Preparation of this manuscript was funded by ACC. The authors declare that they have no other actual or potential competing financial interests.

Acknowledgments

The authors have sole responsibility for the writing and contents of this paper. This paper represents the professional opinions of the authors and not necessarily those of the American Chemistry Council (ACC), which provided funding for this paper.

References

Au, W.W., 2007. Usefulness of biomarkers in population studies: from exposure to susceptibility and to prediction of cancer. Int. J. Hyg. Environ. Health 210 (3–4), 239–246.

Boobis, A.R., Doe, J.E., Heinrich-Hirsch, B., Meek, M.E., Munn, S., Ruchirawat, M., et al., 2008. IPCS framework for analyzing the relevance of a noncancer mode of action for humans. Crit. Rev. Toxicol. 38 (2), 87–96.

Butterworth, B.E., 2006. A classification framework and practical guidance for establishing a mode of action for chemical carcinogenesis. Regul. Toxicol. Pharmacol. 45 (1):9–23. http://dx.doi.org/10.1016/j.yrtph.2006.01.011.

Cox, L.A., Popken, D.A., Kaplan, A.M., Plunkett, L.M., Becker, R.A., 2016. How well can in vitro data predict in vivo effects of chemicals? Rodent carcinogenicity as a case study. Regul. Toxicol. Pharmacol. 77:54–64. http://dx.doi.org/10.1016/j.yrtph.2016.02.005.

European Commission, 2017. ToxRTool - Toxicological Data Reliability Assessment Tool: Instructions for Use. Brussels:Joint Research Centre, Institute for Health and Consumer Protection (3pp).

FAO/WHO, 2004. Toxicological evaluation of paraquat. WHO/PCS/04.1. Pesticide Residues in Food: 2003: Toxicological Evaluations. Joint Meeting of the FAO Panel of Experts on Pesticides Residues in Food and the Environment and the WHO Core Assessment Group. World Health Organization (WHO), Geneva, Switzerland: pp. 203–266 (Available: https://web.archive.org/web/20140904104309/http://whqlibdoc.who.int/publications/2004/924166519X_paraquat.pdf [accessed 18 July 2016]).

Goodman, J.E., Prueitt, R.L., Sax, S.N., Pizzurro, D.M., Lynch, H.N., Zu, K., et al., 2015. Ozone exposure and systemic biomarkers: evaluation of evidence for adverse cardiovascular health impacts. Crit. Rev. Toxicol. 45 (5):412–452. http://dx.doi.org/10.3109/10408444.2015.1031371.

Goodson III, W.H., Lowe, L., Carpenter, D.O., Gilbertson, M., Ali, A.M., de Cerain Salsamendi, A.L., et al., 2015. Assessing the carcinogenic potential of low-dose exposures to chemical mixtures in the environment: the challenge ahead. Carcinogenesis 36 (Suppl. 1): S254–S296. http://dx.doi.org/10.1093/carcin/bgv039.

Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. Cell 144 (5): 646–674. http://dx.doi.org/10.1016/j.cell.2011.02.013.

Hattori, N., Ushijima, T., 2016. Epigenetic impact of infection on carcinogenesis: mechanisms and applications. Genitourin. Med. 8 (1):10. http://dx.doi.org/10.1186/s13073-016-0267-2.

Hill, A.B., 1965. The environment and disease: association or causation? Proc. R. Soc. Med. 58 (5), 295–300.

IARC (International Agency for Research on Cancer), 2015a. IARC Monographs Questions and Answers. International Agency for Research on Cancer, Lyon, France (Available: http://monographs.iarc.fr/ENG/News/Q&A_ENG.pdf [accessed 20 June 2016]).

IARC (International Agency for Research on Cancer), 2015b. Preamble to the IARC Monographs on the Evaluation of Carcinogenic Risks to Humans (Updated). International Agency for Research on Cancer, Lyon, France (Available: http://monographs.iarc.fr/ENG/Preamble/CurrentPreamble.pdf [accessed 15 June 2016]).

IARC (International Agency for Research on Cancer), 2015c. IARC Monographs Volume 112: Evaluation of Five Organophosphate Insecticides and Herbicides. International Agency for Research on Cancer, Lyon, France (Available: https://www.iarc.fr/en/media-centre/iarcnews/pdf/MonographVolume112.pdf [accessed 28 June 2016]).

IARC (International Agency for Research on Cancer), 2016a. Instructions to Authors for the Preparation of Drafts for IARC Monographs. International Agency for Research on Cancer, Lyon, France (Available: https://monographs.iarc.fr/ENG/Preamble/previous/Instructions_to_Authors.pdf [accessed 18 July 2016]).

IARC (International Agency for Research on Cancer), 2016b. Diazinon. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 112: Some Organophosphate Insecticides and Herbicides: Diazinon, Glyphosate, Malathion, Parathion, and Tetrachlorvinphos. International Agency for Research on Cancer, Lyon, France (Available: http://monographs.iarc.fr/ENG/Monographs/vol112/mono112-08.pdf [accessed 9 June 2016]).

Klimisch, H.J., Andreae, M., Tillmann, U., 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul. Toxicol. Pharmacol. 25 (1), 1–5.

Li, N., Parrish, M., Chan, T.K., Yin, L., Rai, P., Yoshiyuki, Y., et al., 2015. Influenza infection induces host DNA damage and dynamic DNA damage responses during tissue regeneration. Cell. Mol. Life Sci. 72 (15):2973–2988. http://dx.doi.org/10.1007/s00018-015-1879-1.

Lynch, H.N., Goodman, J.E., Tabony, J.A., Rhomberg, L.R., 2016. Systematic comparison of study quality criteria. Regul. Toxicol. Pharmacol. 76:187–198. http://dx.doi.org/10.1016/j.yrtph.2015.12.017.

Lynch, H.N., Loftus, C.T., Cohen, J.M., Kerper, L.E., Kennedy, E.M., Goodman, J.E., 2017. Weight-of-evidence evaluation of associations between particulate matter exposure and biomarkers of lung cancer. Regul. Toxicol. Pharmacol. (In press).

Meek, M.E., Bucher, J.R., Cohen, S.M., Dellarco, V., Hill, R.N., Lehman-McKeeman, L.D., et al., 2003. A framework for human relevance analysis of information on carcinogenic modes of action. Crit. Rev. Toxicol. 33 (6), 591–653.

Meek, M.E., Boobis, A., Cote, I., Dellarco, V., Fotakis, G., Munn, S., et al., 2014a. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J. Appl. Toxicol. 34 (1):1–18. http://dx.doi.org/10.1002/jat.2949.

Meek, M.E., Palermo, C.M., Bachman, A.N., North, C.M., Lewis, R.J., 2014b. Mode of action human relevance (species concordance) framework: evolution of the Bradford Hill considerations and comparative analysis of weight of evidence. J. Appl. Toxicol. http://dx.doi.org/10.1002/jat.2984.

Miller, M.F., Goodson III, W.H., Manjili, M.H., Kleinstreuer, N., Bisson, W.H., Lowe, L., 2016. Low-dose mixture hypothesis of carcinogenesis workshop: scientific underpinnings and research recommendations. Environ. Health Perspect. http://dx.doi.org/10.1289/EHP411.

Morita, T., Hayashi, M., Nakajima, M., Tanaka, N., Tweats, D.J., Morikawa, K., et al., 2009. Practical issues on the application of the GHS classification criteria for germ cell mutagens. Regul. Toxicol. Pharmacol. 55:52–68. http://dx.doi.org/10.1016/j.yrtph.2009.06.001.

Pilger, A., Rudiger, H.W., 2006. 8-hydroxy-2′-deoxyguanosine as a marker of oxidative DNA damage related to occupational and environmental exposures. Int. Arch. Occup. Environ. Health 80 (1), 1–15.

Pilger, A., Germadnik, D., Riedel, K., Meger-Kossien, I., Scherer, G., Rudiger, H.W., 2001. Longitudinal study of urinary 8-hydroxy-2′-deoxyguanosine excretion in healthy adults. Free Radic. Res. 35 (3):273–280. http://dx.doi.org/10.1080/10715760100300811.

Preston, R.J., Hoffman, G.R., 2013. Genetic toxicology. In: Klaassen, C.D. (Ed.), Casarett and Doull's Toxicology: The Basic Science of Poisons, eighth ed. McGraw-Hill Education, New York, pp. 445–480.

Rawlings-Goss, R.A., Campbell, M.C., Tishkoff, S.A., 2014. Global population-specific variation in miRNA associated with cancer risk and clinical biomarkers. BMC Med. Genet. 7 (53). http://dx.doi.org/10.1186/1755-8794-7-53.

Samuel, G.O., Hoffmann, S., Wright, R.A., Lalu, M.M., Patlewicz, G., Becker, R.A., et al., 2016. Guidance on assessing the methodological and reporting quality of toxicologically relevant studies: a scoping review. Environ. Int. 92–93:630–646. http://dx.doi.org/10.1016/j.envint.2016.03.010.

Schneider, K., Schwarz, M., Burkholder, I., Kopp-Schneider, A., Edler, L., Kinsner-Ovaskainen, A., et al., 2009. 'ToxRTool', a new tool to assess the reliability of toxicological data. Toxicol. Lett. 189 (2), 138–144.

Smith, M.T., Guyton, K.Z., Gibbons, C.F., Fritz, J.M., Portier, C.J., Rusyn, I., et al., 2016. Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. Environ. Health Perspect. 124 (6):713–721. http://dx.doi.org/10.1289/ehp.1509912.

Stephens, M.L., Betts, K., Beck, N.B., Cogliano, V., Dickersin, K., Fitzpatrick, S., et al., 2016. The emergence of systematic review in toxicology. Toxicol. Sci. 152 (1):10–16. http://dx.doi.org/10.1093/toxsci/kfw059.

Villeneuve, D.L., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T.H., LaLone, C.A., et al., 2014. Adverse outcome pathway (AOP) development I: strategies and principles. Toxicol. Sci. 142 (2):312–320. http://dx.doi.org/10.1093/toxsci/kfu199.

Willhite, C.C., 2001. Weight-of-evidence versus strength-of-evidence in toxicologic hazard identification: di(2-ethylhexyl)phthalate (DEHP). Toxicology 160 (1–3): 219–226. http://dx.doi.org/10.1016/S0300-483X(00)00451-0.